



Visualiser la complexité du droit

Danièle Bourcier^(a), Pierre Mazzega^(b) et Romain Boulet^(b)

^(a) CERSA UMR7106, CNRS Université Paris II

^(b) LMTG UMR5563, CNRS, IRD, Université Toulouse III

1 Notre problématique

Peu de recherches en théorie juridique ont été faites sur la complexité normative¹ et encore moins, à notre connaissance, sur ce que nous voulons explorer et visualiser à l'intérieur d'un corpus juridique : les lieux (et non les causes) de la complexité du droit. Nous avons limité notre analyse au corpus codifié pour tester la pertinence de nos outils de représentation. Codifier est en effet une opération (particulière) d'organisation des textes et des normes destinée à présenter la connaissance juridique sous une certaine *uniformité* face « à un amas confus et informe, un immense chaos » - comme le déplorait déjà un des rédacteurs prestigieux du Code Civil, Portalis.

La codification donne lieu à une réorganisation de la connaissance en fonction du passé mais aussi d'une certaine vision du futur et des évolutions possibles. La structuration de cette forme est contrainte par le matériau existant mais, comme un sculpteur, le codificateur possède une certaine marge de liberté dans l'organisation de cette connaissance. La codification crée des réseaux de références entre normes à l'intérieur du code et en direction de textes extérieurs.

Pourtant, lorsqu'on consulte un code, l'impression générale d'*uniformité* laisse place à l'apparition d'une multitude de niveaux et de disparités internes dans l'organisation des articles. En effet, le code a vocation à évoluer dynamiquement sous l'effet de nouvelles lois ou décrets qui viennent modifier, augmenter, annuler les textes codifiés: cette belle architecture initiale évolue donc sous l'effet des évolutions normatives qui transforment une certaine organisation interne de la connaissance. La représentation des phénomènes qui avaient contribué à réorganiser cette structure puis à la reconfigurer dynamiquement peut être systématisée en s'appuyant sur la théorie des graphes².

Dans un premier temps, organiser cette représentation avec des outils est un moyen d'observer, de quantifier et de visualiser ces faits. Notre article propose pour développer ces hypothèses d'analyser deux niveaux d'organisation de corpus juridiques : a) un niveau macro-organisationnel, à savoir le réseau des inter-citations entre tous les codes en vigueur en Droit Interne (France) ; b) un niveau micro-organisationnel : la structure de la partie législative du Code de l'Environnement (notée *CEL* par la suite). Précisons dès à présent que le Code de

¹ P.H. Schuk, 'Legal complexity : some causes, consequences, and cures', *Duke Law Journal*, 42 (1), 1-52, 1992; voir aussi une autre approche, celle des économistes du droit : M. White, 'Legal complexity and Lawyer's benefit from litigation international', *Rev. of Law and Economics*, vol. 12, 381- 383, n° 5, 1992; E MacKaay 'Les notions floues en droit ou l'économie de l'imprécision', N° Spécial sur le discours juridique, Danièle Bourcier (éd.) *Revue Langages*, (1979) n°53, pp 33-50.

² Dans cet article, nous n'entrons pas à proprement parler dans la théorie des graphes mais présentons des résultats d'analyse que nous avons obtenus en utilisant des concepts et méthodes issus de cette théorie.

l'environnement a hérité des particularités du droit de l'environnement, droit très jeune, au carrefour de plusieurs autres matières, et qu'il est souvent considéré comme un « droit patchwork »³.

Après avoir décrit les caractéristiques générales de la structure de la connaissance juridique (section 2), nous établissons le réseau des codes internes et montrons l'existence de communautés entre codes que l'analyse mathématique permet de mettre en évidence (section 3). Nous entrons ensuite à l'intérieur d'un code pour en analyser la structure et nous interroger sur la pertinence de la notion classique d'article juridique dans le contexte d'une évolution des systèmes juridiques vers plus de complexité (section 4). En effet l'article est de façon canonique l'unité de base du code en tant qu'il se situe à l'articulation d'une vision structurante et d'une vision sémantique de la connaissance. Nous constaterons que la notion d'article a changé et que coexiste, à côté de cette vision unitaire, *une vision plus fragmentée et dynamique de la norme textuelle*. Nous développerons cette hypothèse en utilisant les résultats d'un modèle de description de nos données sous forme de graphes pour en conclure que le législateur, le codificateur et le juriste auraient intérêt à intégrer un niveau plus pertinent, celui de document numérique virtuel, pour traiter cognitivement la complexité des textes juridiques (section 5). Les perspectives de cette analyse sont abordées en conclusion (section 6).

2 Caractériser la connaissance juridique

La connaissance juridique est composée de l'ensemble des textes juridiques qui sont nécessaires à l'application ou à l'interprétation d'une situation juridique. Cette connaissance a des caractéristiques précises.

2.1 La connaissance juridique est interconnectée

La connaissance juridique est composée de textes à portée normative (la loi, le décret, l'arrêté)⁴. Les textes qui composent cette connaissance sont *rassemblés* de façon particulière. Ainsi aucun texte nouveau ne peut être inclus dans le périmètre de cette connaissance sans qu'il soit fait état de ses liens avec le droit existant. C'est certainement la caractéristique la plus saillante de cette connaissance: elle n'est pas accumulative, elle est *constitutivement* interconnectée. La philosophie du droit a systématisé ces formes constitutives avec une hiérarchie (Kelsen) suivant laquelle les règles inférieures doivent respecter les règles supérieures. Cette interconnexion à la source ne signifie pas qu'un nouveau texte fait renvoi à tous les autres textes mais que le Législateur doit tenir compte de l'ensemble du droit pour inclure ces nouvelles dispositions dans le monde juridique.

Pour parer le jeu des effets directs ou secondaires de l'introduction d'une nouvelle norme, le législateur doit anticiper les dispositions qui ne seraient plus applicables compte tenu du nouveau texte ou en créer d'autres qui devraient les remplacer. Il lui arrive parfois d'oublier une liaison qui aurait dû être faite et cet oubli peut entraîner des conflits d'interprétation, ou au pire des *effets pervers* qui sont connus notamment dans le domaine du droit fiscal. Il lui arrive aussi de ne pas prévoir tous les effets d'un nouveau texte. Par la suite, l'interprète du droit, qu'il soit juge ou administrateur, devra restituer ces liens pour interpréter la règle en question par rapport à l'ensemble du droit.

³ P. Lascoumes & G. Martin, « Des Droits épars au code de l'environnement », *Droit & Société*, n° 30-31, 1995.

⁴ Les circulaires par exemple ne sont pas normatives mais décrivent la façon dont les dispositions normatives peuvent être interprétées.

Cette hypothèse forte de la fonction intégrative du système juridique explique pourquoi le législateur indique non seulement la règle qui devra être modifiée par le texte nouveau mais intervient aussi dans le futur travail de l'interprète en lui signalant comment ces nouvelles règles peuvent influencer des interprétations issues de ce nouveau texte.

La forte intégration de la connaissance juridique conduit à l'observation du phénomène suivant: les textes juridiques sont fortement inter-reliés, car ils portent au fur et à mesure de l'évolution de la législation les traces de cette interconnexion substantielle⁵. La nature des liens qui relient les textes est double : les liens sont structurels et sémantiques⁶. Les liens structurels sont des liens explicites à l'organisation interne ou externe des textes. Dans un code, les liens entre les dispositions se tissent à l'intérieur du même code ou en référence à des dispositions d'autres codes et lois.⁷ Les liens sémantiques portent sur les contenus ontologiques et ne sont pas systématiquement explicités car ils se « réfractent » dans le droit.

2.2 La connaissance juridique est sécable

Le texte juridique (loi, code jusqu'à l'arrêté) est divisé en articles. A partir de la circulaire, le texte peut ne pas être divisé en articles mais en sections, ou même organisé sans indication particulière. On proposera une première explication de la « divisibilité » de cette connaissance : l'organisation en articles permet de faire des citations ultérieures au niveau de l'article. L'article est caractérisé par un numéro. Il n'a pas de nom, il n'a pas de titre. Son numéro indique sa place précise dans l'édifice des normes. L'article est la brique fondamentale du système juridique⁸ :

- 1- L'article est avec la loi le seul élément que peut se remémorer le juriste (par ex. le fameux « Article 1382 du code civil ») et qui peut être cité comme une référence fondamentale (par ex. la « loi Barnier »).
- 2- L'article est contenu dans un plan hiérarchiquement organisé. Mais il n'est pas la plus petite partie sur laquelle porte une modification du législateur : en dessous de lui, l'alinéa, la phrase et les mots sont encore des unités plus microscopiques sur lesquelles peut porter l'évolution du droit.
- 3- L'article est au croisement du plan (hiérarchique, compréhensif ou extensif) et du réseau des références, deux axes qui constituent l'organisation fondamentale de la connaissance juridique.
- 4- L'article a aussi une dimension syntagmatique dans la linéarité du discours juridique. Il est un élément dans la continuité d'un groupe d'articles qui relève d'un même niveau dans le plan. Cependant deux articles qui se suivent peuvent avoir des univers de références différents puisqu'il peut exister une rupture au niveau du plan⁹.

A cause de ces différents liens (syntagmatiques, arborescents et hypertextuels), l'article de loi est lesté d'un certain nombre d'appels à références, ou de citations complètes au-delà de l'information particulière qu'il contient.

⁵ L'estimation de la densité du réseau des liens entre codes que nous verrons en section 3 donne en quelque sorte une mesure d'un effet de cette propriété à l'échelle du système du droit interne.

⁶ D. Bourcier & P. Mazzega (2007) "Toward measures of legal complexity", *Proc. 11th Intern. Conf. on Artificial Intelligence and Law*, Stanford Law School, ACM Press, New York, 211-215.

⁷ Ce système de références est caractéristique du monde de l'écrit qui fige un état matériel de la connaissance. Dans le monde de l'oral, la référence explicite entre les connaissances n'est pas aussi précise. C'est la linéarité du temps dans le récit plus que la référence à la structure de la connaissance qui reliera les interprétations car l'oral suit d'autres contraintes cognitives de mémorisation.

⁸ Cette partition en articles est ancienne : les Tables de la loi, le code Hammourabi sont aussi divisés en articles.

⁹ Par exemple deux articles contigus dans le texte mais appartenant à deux chapitres distincts.

Dans le guide de légistique de la banque de données juridiques officielle LEGIFRANCE¹⁰, on définit l'article comme étant « l'unité de base d'un texte normatif », ce qui signifie qu'il est préférable qu'une seule règle soit énoncée par article. Il est même précisé qu'il existe une longueur optimale d'article permettant d'éviter les articles trop longs qui devront comporter de nombreuses subdivisions. L'article organise le droit. Même au moment du vote, c'est le niveau de l'article qui est mis aux voix. Ainsi si l'article est clair et cohérent, le débat et l'expression du vote en sont facilités. Il en est de même de l'interprétation et des modifications ultérieures.

L'article est une unité normative de sens : il est relativement autonome. Comme la numérotation des articles est linéaire, la lecture des articles peut être linéairement cohérente et discursive : le sens se déploie au fur et à mesure de la lecture. Autonome, l'article a aussi une identité : cette identité se manifeste au niveau de l'intangibilité du numéro. Le Guide demande que si possible on le maintienne, pour éviter que, lorsqu'on modifie un code, on ne tienne pas compte des références à des articles qui peuvent figurer dans d'autres textes. Les articles nouveaux devront être insérés en tenant compte de cette intangibilité. Dès que l'information de référence nécessite de dépasser plusieurs articles, les numéros deviendront essentiels pour définir les dispositions concernées.

Quand il s'agit de faire appel aux *liens hiérarchiques* comme ceux du plan ou de la table des matières (particulièrement importants quand il s'agit de codes), une autre forme de référence doit être utilisée pour lier les articles à cette structure. Cet appel ne sera pas révélé au sein d'une image mentale linéaire de la connaissance mais d'une arborescence : l'article ou la suite d'articles indiqués par leur numéro, ou la partie de la table des matières indiquée par son niveau, son titre, voire ses articles seront cités.

Quand il s'agit de faire appel aux *liens réticulaires*, les références se feront essentiellement par le numéro d'article ou de groupe d'articles. Dans ce cas c'est le contenu lui-même qui est « référé » et non la structure. Le législateur a le choix pour expliciter la nature du lien sémantique en paraphrasant ou en reprenant intégralement le morceau de texte cité. Les références peuvent être croisées mais ce cas reste assez rare : c'est pour cela que la plupart des bases de codification parlent de références *fantômes* pour expliciter l'ensemble des liens qui unissent les éléments de la connaissance.

Enfin, d'autres liens peuvent être considérés : les liens de codes à codes. La codification en France est une réécriture du droit à partir de textes existants. Une structure interne des codes s'exprime sous forme de plan ; mais il existe une hyper-structure des codes¹¹. On en trouve la trace dans la délimitation du périmètre d'un code par rapport à un autre : par exemple quels textes feront partie du code rural ou du code de l'environnement¹². Un effet plus visible de cette hyper-structure se situe dans le débat sur les codes pilotes ou les codes suiveurs.

2.3 Norme et document juridiques

Les termes de « norme », « article », « document » ont été utilisés dans des contextes qui doivent être explicités. Deux dimensions doivent être prises en compte : la dimension textuelle et la dimension substantielle.

La norme peut être considérée comme l'unité *juridique* de base du droit : le droit est un univers de normes. Mais le législateur ne fait jamais référence à la notion abstraite de norme. La norme est une construction substantielle élaborée soit *a priori* (avant sa formulation sous

¹⁰ Disponible sur le site LEGIFRANCE : <http://www.legifrance.gouv.fr/>.

¹¹ C'est la Commission supérieure de la codification qui décide de la planification de la codification.

¹² Délimitation qui fait l'objet parfois de longs arbitrages tant le code avec son nom peut apparaître comme une conquête institutionnelle de première importance entre les ministères.

forme de texte), soit *a posteriori* pour analyser le texte voire le mettre sous forme de règle de production. Ainsi dans un article l'opération qui consiste à traduire des textes sous forme de règles de production (normes et conditions de la norme) a été décrite comme une opération de représentation de connaissances. Qu'elle soit élaborée par des philosophes du droit, par des logiciens ou des cognitivistes, la norme n'existe pas à l'état pur.

L'article a l'intérêt de pouvoir être considéré comme unité formelle et comme unité substantielle. Pour Portalis, qui est le père du Code civil et son méthodologue, l'article devait être une unité de sens. Aujourd'hui l'article est devenu plus flou pour ce qui est de la dimension « substantielle » ; en revanche sa forme est aisément définissable.

Le document appartient à un autre champ fonctionnel: le document est sans nul doute venu de l'expression américaine « document » notamment dans les premiers logiciels documentaires comme le DPS (*Document Processing System*). Rappelons que ce logiciel avait été créé au départ pour traiter des documents judiciaires dans un dossier de contentieux. Ce n'est que par la suite qu'une application documentaire au texte juridique a été conçue.

Pour le droit, le document peut être un article, une section (circulaire), une loi. Le document est défini par le système d'information qui le rend opérationnel¹³.

Quand on parle de document juridique complexe, on veut tenir compte de cette notion de document, unité documentaire des banques de données juridiques ; mais on veut aussi y ajouter une dimension particulière que nous décrivons par rapport à une nouvelle approche qui considère le droit comme un système complexe.

Depuis quelques décennies, les usagers du droit se plaignent de *la complexité du droit*. Nous avons pris le parti de dire que si cette complexité ne peut être réduite, elle peut être gérée, traitée, maîtrisée. L'approche d'un système par sa complexité nous permet de développer les outils d'analyse scientifique et technique de cette complexité, les algorithmes de son usage. On peut simplifier un système compliqué. On ne peut en revanche réduire un système complexe dont l'évolution résulte généralement d'un processus d'auto-organisation non supervisé de nombreuses entités. Un système complexe est le résultat d'enchevêtrements et d'interactions à plusieurs niveaux évoluant dynamiquement.

Le document complexe est un niveau de traitement de la complexité. La représentation des graphes et leur analyse forment une approche susceptible de prendre en compte les liens (structuraux et sémantiques) que nous avons décrits précédemment. Même si ces liens sont pléthore (plusieurs milliers pour la seule partie législative du code de l'environnement - *CEL*), leur nombre n'est pas un inconvénient mais une richesse d'informations qu'il faut prendre en compte. Ainsi de nouvelles informations seront susceptibles de considérer le document sous divers points de vue.

3 Le réseau des codes de droit français

La politique de codification mise en place en 1989 a conduit à codifier 70 % des textes isolés du droit interne. Nous proposons de produire une forme de visualisation de la structure de l'ensemble de cet imposant édifice juridique. Ce projet est en soi inhabituel, la connaissance du droit étant généralement limitée à des domaines aux contours plus délimités, ou à des branches du droit. Comme il a été décrit ci-dessus, l'acte de légiférer est réalisé dans un contexte de connaissance partielle de l'édifice juridique existant. En d'autres termes, la production de textes normatifs n'est pas soumise à un contrôle centralisé ayant en vue

¹³ Le document est défini dans le système DOCILIS (premier système documentaire du Centre d'Etudes et d'Information Juridiques CEDIJ, d'où est né LEGIFRANCE) comme « unité de gestion et de recherche documentaire ».

l'ensemble de la structure et de ses contenus. Mais la vue d'ensemble fait défaut au codificateur : elle peut désormais être visualisée.

3.1 Le graphe représentant le réseau des codes

Parmi les codes en vigueur disponibles sur le site officiel LEGIFRANCE, nous n'avons retenu que les 52 codes qui ont été produits sous la responsabilité de la Commission Supérieure de Codification (CSC) selon les méthodes décrites dans le guide de légistique (*op. cit.*). Parmi les codes non retenus, le plus remarquable est sans doute le Code Général des Impôts qui a été conçu par le Ministère du Budget. Nous avons ensuite compté le nombre de fois où un code A citait un article ou une autre subdivision (chapitre, titre, livre ou code lui-même) d'un code B. Pour représenter le réseau des citations entre codes nous procédons ainsi : nous construisons un graphe dont chaque sommet est un code – par exemple les sommets libellés A et B en référence aux codes évoqués ci-dessus, un lien étant tracé entre les sommets A et B si un des deux codes cite au moins une fois l'autre.

Cette méthode appelle quelques commentaires. Le graphe que nous construisons pour représenter le réseau des citations entre codes n'est pas orienté : nous ne retenons pas l'information qui indique si A cite B ou à l'inverse si c'est B qui cite A. Le graphe n'est pas non plus pondéré : que le code A cite une seule fois ou de nombreuses fois le code B est traité de la même manière, en ne représentant qu'un lien qui ne comporte pas l'information du nombre de citations (ce nombre étant une sorte de poids attribuable au lien). D'autre part, les sommets sont ici des codes, sans considération pour leurs divisions internes (notamment sans distinction des articles qui les constituent). Ainsi le graphe associé au réseau des codes se situe à un niveau de résolution défini – celui où les seuls objets reconnus sont des codes – et ne conserve que l'information qui indique si deux codes (sommets) sont liés ou non¹⁴.

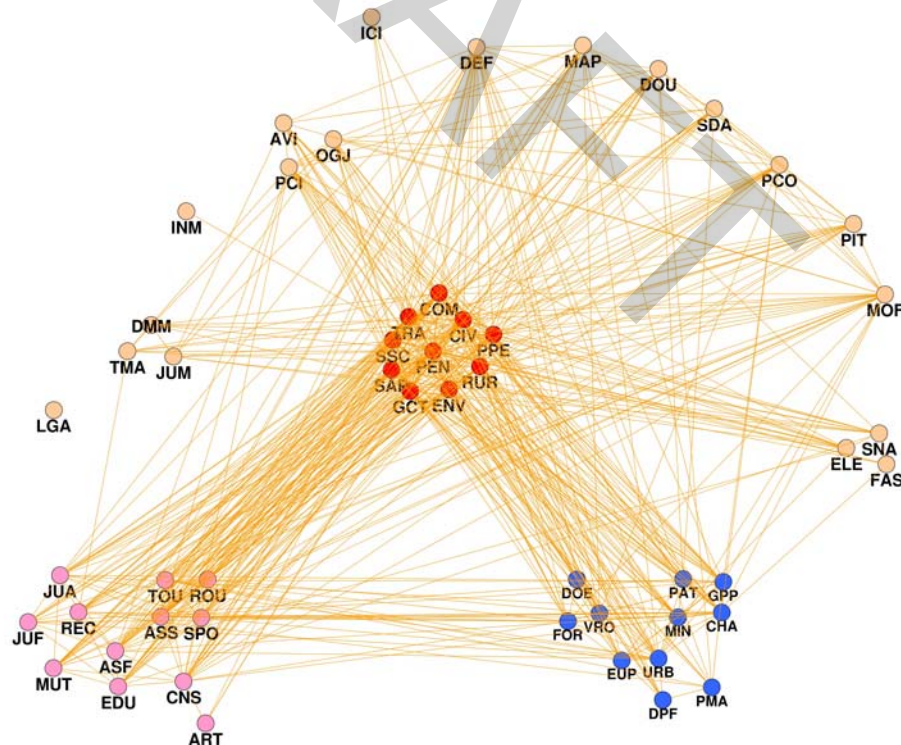


FIGURE 1 : Graphe représentant le réseau des codes français, avec regroupement partiel en communautés. Au centre le club huppé (CIV : code civil, COM : commerce, ENV : environnement, GCT : général des

¹⁴ Mazzega P., Bourcier D. and R. Boulet (2009a) *The network of French legal codes*. Proc. 12th Conf. on Artificial Intelligence and the Law ICAIL, Barcelona, ACM Press, 236-237.

collectivités territoriales, PEN : pénal, PPE : procédure pénale, RUR : rural, SAP : santé publique, SSC : sécurité sociale, TRA : travail). En bas à gauche, une communauté stable de 12 codes (ART : artisanat, ASF : action sociale et familles, ASS : assurances, CNS : consommation, EDU : éducation, JUA : justice administrative, JUF : juridictions financières, MUT : mutualité, REC : recherche, ROU : route, SPO : sport, TOU : tourisme) ; en bas à droite, une communauté stable de 11 codes (CHA : construction et habitation, DOE : domaine de l'Etat, DPF : domaine public fluvial, EUP : expropriation utilité publique, FOR : forestier, GPP : général de la propriété des personnes publiques, MIN : minier, PAT : patrimoine, PMA : ports maritimes, URB : urbanisme, VRO : voirie routière) ; les autres codes ont un positionnement plus ou moins stable suivant les critères de regroupements utilisés (de droite à gauche : MOF : monétaire et financier, PIT : propriété intellectuelle, PCO : postes et communications électroniques, SDA : entrée séjour et droit d'asile, DOU : douanes, MAP : marchés publics, DEF : défense, ICI : industrie cinématographique, AVI : aviation civile, OGJ : organisation judiciaire, PCI : procédure civile), avec deux petits regroupements (DMM : disciplinaire et pénale marine marchande, TMA : travail maritime, JUM : justice militaire ; et d'autre part ELE : électoral, SNA : service national, FAS : famille et aide sociale). Le code INM (instruments monétaires et médailles) est pendant (cite 1 code mais non cité) et le LGA (légion d'honneur, médailles militaires) est isolé du reste du graphe.

Notons dès à présent qu'une analyse préalable du graphe orienté pondéré (tenant compte du sens des citations et de leur nombre) conduit à des conclusions comparables à celles que nous présentons ici. Ce résultat semble signifier que le fait que des codes soient liés deux à deux est aussi structurant pour l'édifice complet que le fait que certains codes ne soient pas liés, créant ainsi une sorte de carte des dépendances entre contenus ou substances des codes respectifs. En effet notre analyse porte principalement sur la recherche de communautés entre groupes de codes, communautés dont les contours (liste des membres constituants) s'avèrent robustes. Cette robustesse est également vérifiée en soumettant l'analyse du graphe à divers algorithmes de partitionnement en communautés. Ces algorithmes reposent sur différents critères de regroupement des codes dont nous ne présenterons pas ici les concepts mathématiques ni la traduction algorithmique¹⁵, pour insister d'avantage sur les résultats obtenus.

Le graphe représentant le réseau des 52 codes retenus est présenté en Figure 1. La première caractéristique de ce graphe est sa forte densité. La densité d'un graphe est le rapport du nombre de liens effectifs divisé par le nombre possible de liens entre tous les sommets pris deux à deux. Dans les réseaux sociaux¹⁶, cette densité est de l'ordre de quelques pourcents. Le graphe des codes, avec une densité de près de 40%, fait figure d'exception, et exprime la forte intégration de la connaissance juridique déjà évoquée (voir sec.2.1). Nous suspectons d'ailleurs cette forte densité d'être une propriété typique de nombreux réseaux de textes juridiques (pas nécessairement français), au moins considérés à cette échelle de résolution (c'est-à-dire où les sommets seraient des codes ou des lois).

¹⁵ Le lecteur intéressé consultera Boulet R., Mazzega P. et D. Bourcier (2009) *Analyse d'un graphe juridique dense : le monde concentré des codes législatifs*. Techniques & Sci. Informatiques, n° spécial *Graphes de Terrain*, M. Latapy ed., soumis.

¹⁶ Il existe une vaste littérature sur l'analyse de « réseaux sociaux » tels que le réseau des acteurs de cinéma (Newman M. E. J. (2000) *Models of the small world*. J. Stat. Phys., vol.101, 819-841), le réseau des collaborations scientifiques (Grossman J. (2002) *The evolution of the mathematical collaboration graph*. Congressus Numerantium, vol.158, 202-212), le réseau des relations amoureuses au sein d'un lycée (Beaman P., Moody J. and K. Stovel (2004) *Chains of affection : the structure of adolescent romantic and sexual networks*. Am. J. Sociology, vol.100, 44-91) ou le réseau social entre paysans au Moyen Age (Boulet R., Jouve B., Rossi F. et N. Villa (2008) *Batch kernel SOM and related Laplacian methods for social network analysis*. Neurocomputing, vol.71, 1257-1273).

3.2 Le club huppé des codes

Dix codes présentent un fort degré de connectivité¹⁷ (nombre de liens important : 66 pour le code général des collectivités territoriales, 64 pour le code de la santé publique, 62 pour le code de procédure pénale, etc.). Le code pénal est le plus cité. On constate également qu'ils sont tous liés entre eux deux à deux (à une exception près), et que chacun d'eux est lié à la plupart des 42 autres codes. Cette structure locale de forte densité interne (voisine de 100%) et de forte centralité est appelée « club huppé » (*rich club* en anglais) et se retrouve dans certains réseaux sociaux (club des individus les plus influents). Ces dix codes (cf. Fig.1) forment en quelque sorte la base incontournable du droit interne actuel. L'appartenance de chacun de ces codes au club huppé résulte probablement de raisons distinctes : le code civil est historiquement antérieur aux autres codes et à ce titre est souvent cité ; toutes les formes de pénalités et de sanctions, quelque soit le domaine juridique concerné, sont rassemblées dans le code pénal ; le code de l'environnement touche une très grande diversité de domaines eux-mêmes complexes (santé, ressources, territoires, gouvernance, etc.) ; *idem* pour les codes de la santé publique, du commerce ou de la sécurité sociale, etc. Une analyse plus fine des relations au sein du club huppé consistera à prendre en compte les orientations et les poids attribués aux liens entre codes membres.

3.3 Autres communautés de codes

L'application de divers critères mesurant la connectivité entre codes et la densité relative de leurs liens au sein de regroupements divers permet, une fois retiré les codes du club huppé, de mettre en évidence des communautés cachées. Au sein d'une même communauté, la densité de liens entre membres est nettement plus élevée que la densité globale du réseau. L'application de critères mathématiques (certains ayant été développés pour l'analyse de réseaux sociaux) permet de s'affranchir d'une éventuelle intuition juridique, et en cela présente un caractère que nous pourrions qualifier d'objectif. L'usage de plusieurs critères (et l'application des algorithmes associés) permet en outre de tester la robustesse de l'identification d'une communauté. Nous ne parlons ici que de communautés robustes, stables vis-à-vis de l'usage de plusieurs critères d'appartenance à une communauté. Se pose ensuite la question de l'interprétation ou de la raison de tels regroupements. Par leur nature ces communautés n'ont pas été jusqu'à présent repérées. Elles ont été construites, en quelque sorte en aveugle, au cours de l'évolution de l'ensemble de l'édifice du droit interne et de sa codification. Pour les mêmes raisons, bien que l'identité de ces communautés ne suive pas les divisions disciplinaires habituelles, leur existence est indéniable.

La communauté de 12 membres représentée en bas à gauche de la Figure 1, rassemble des codes régulant des domaines d'activités collectives à caractère de service public: les actions sociales au sens large, la consommation et l'éducation, le sport et le tourisme, les assurances et la mutualité, etc. En bas à droite la communauté stable regroupe 11 codes régulant des domaines touchant les territoires, les ressources naturelles, en bref l'environnement. Il n'est probablement pas fortuit que les codes traitant de la propriété ou du patrimoine appartiennent également à cette communauté. L'identification des raisons sous-jacentes, ou des liens effectifs qui rassemblent ces codes dans ces communautés auparavant inconnues requiert une analyse plus fine de leurs contenus, de leurs substances et de leurs interdépendances cognitives.

Enfin les autres codes ne constituent pas à proprement parler une communauté large et stable. Une explication partielle probable est que certains des codes pourraient appartenir à

¹⁷ Voir Mazzega P., Bourcier D. and R. Boulet (2009b) *Code communities in the French legal system*, Proc. Workshop *Privacy & Protection in Web-based Social Networks*, ICAIL Barcelona, submitted.

plusieurs communautés, de manière similaire à ce qui peut être observé de certains individus liés simultanément à plusieurs communautés.

D'autres domaines du droit se présentent principalement, voire uniquement, sous forme de textes liés au travers de réseaux - « maillage de conventions internationales, bilatérales et multilatérales, qui relie notre pays au reste du monde »¹⁸. Ainsi, « le droit de l'environnement (qui) notamment au niveau international, se constitue progressivement de textes, de conventions organisées en réseau »¹⁹. En distinguant ce type particulier de liens entre textes, notre approche permet d'aborder également toute matière juridique textuelle organisée selon ce seul principe.

4 La structure du code de l'environnement

Nous abordons maintenant l'analyse de la structure d'un code. On ne s'intéresse ici qu'à la structure qui s'ordonne en superposant une organisation hiérarchique (arbre) et un réseau de référencements vers des textes internes ou externes au code considéré²⁰. Cette structure est une des manifestations de la complexité du droit qui s'offre à l'analyse²¹. Elle est mise en évidence en considérant les douze niveaux hiérarchiques suivants: code, partie, livre, titre, chapitre, section, sous section, paragraphe, article, alinéa « romain » et nota, alinéa arabe et alinéas non numérotés. Le corpus analysé pour cette étude comprend la partie législative du Code de l'Environnement²² (*CEL*) ainsi que les textes s'y trouvant directement cités ou référencés (textes « voisins » au sens des graphes, cf. ci-dessous)²³.

4.1 Organisation hiérarchique du code

Chaque niveau est de fait une classe d'objets. Dans notre corpus, le niveau le plus élevé correspond à la classe des « codes ». Le Code de l'Environnement est un objet de cette classe, ainsi que les autres codes cités (dans sa version de 2006 le *CEL* cite 34 autres codes). Par ordre de résolution croissante, nous avons ensuite les niveaux inférieurs suivants: partie (législative ou réglementaire), livre, titre, chapitre, section, sous-section, paragraphe et article. Mais le *CEL* fait aussi référence à des objets juridiques du droit interne, communautaire ou international qui lui sont extérieurs: conventions internationales, ordonnances, décrets, directives, conventions, etc. Tous ces objets seront indistinctement rangés dans la classe

¹⁸ Conseil d'Etat (2006) Rapport public 2006 – Sécurité juridique et complexité du droit, La Documentation française, Etudes et Documents n°57, p.233.

¹⁹ Labrot D. (2007) « Droit et complexité – Regard sur le droit de l'environnement », *Droit et Complexité – Pour une nouvelle intelligence du droit vivant*, Doat, Le Goff, Pédrot (dirs), Presse Univ. Rennes, 17-37.

²⁰ Il serait par exemple possible d'analyser aussi la structure informative qui lie une ontologie juridique et un corpus textuel. Voir Mazzega P., Bourcier D., Bourguin P., Nadah N. and R. Boulet (2009) *A complex-system approach: legal knowledge, ontology, information and networks*. in *Approaches to Legal Ontologies*, Sartor, Casanovas, Biasiotti & Fernandez Narrera eds., Springer, submitted.

²¹ Bourcier D. and P. Mazzega (2007) Toward measures of legal complexity. Proc. 11th Intern. Conf. On Artificial Intelligence and Law ICAIL, Stanford Law School, ACM Press, New York, 211-215.

²² Notre approche est empirique: nous partons du texte tel qu'il est publié au mois de mars 2007 et accessible sur le site LEGIFRANCE à cette date.

²³ Le droit de l'environnement *sensu stricto* est très récent en France. En effet, la partie législative du code a été approuvée par l'Ordonnance n° 2000-914 du 18 septembre 2000 relative à la partie législative du code de l'environnement, ratifiée par la loi n° 2003-591 du 2 juillet 2003 habilitant le gouvernement à simplifier le droit. Les six livres ainsi créés ont été complétés par la loi n° 2003-346 du 15 avril 2003 qui a institué le livre VII consacré à la protection de l'environnement en Antarctique.

“objets externes”²⁴. Une hiérarchie plus ou moins contraignante organise également la matière à l’intérieur même des articles. Nous ne distinguerons que trois niveaux intra-article : les paragraphes dotés d’un chiffre romain (ex.: I, II, etc., auxquels nous ajouterons les « nota » lors de comptages d’objets) sous la dépendance desquels se trouvent les paragraphes dotés d’un chiffre arabe (ex.: 1°, 2°, etc.), et enfin les simples alinéas. Lors des comptages ces objets seront rassemblés ci-dessous sous la classe « intra-article ».

Nous devons rendre compte de deux types de liens entre objets d’un corpus. D’abord le lien qui relie un objet référencé (cible de renvoi, citation, etc.) à l’objet qui le cite (source). C’est ce type de lien qui induit les réseaux mentionnés précédemment. De manière indépendante, la codification organise un corpus rassemblé autour d’une matière (e. g. l’environnement²⁵) elle-même subdivisée (e. g. dans le *CEL*, milieux physiques, espaces naturels, faune et flore, ou encore dispositions applicables à des territoires spécifiques, etc.), ainsi de suite. Cette organisation de la *matière* se traduit par une hiérarchie d’objets dont les articles véhiculent une part de contenu²⁶.

En conséquence nous distinguons les liens (donc les arêtes du graphe) induits par deux types de relations : les relations d’influences et les relations de sélection. La relation d’influence R_I relie un article à l’objet de niveau supérieur auquel il est subordonné. Ainsi, par exemple, dans la Figure 2, l’article L131-8 (en Figure 2, cet article est numéroté 8) est lié par influence à la section 2 (chap. I, titre III, livre I) qui est le niveau supérieur dont il dépend immédiatement (nous dirons plus loin que le chemin entre les sommets correspondants au L131-8 et la section 2 est de longueur 1) alors qu’aucune sous-section ni paragraphe n’existent sous cette section. Du fait de la linéarité du langage (et de l’écriture sur un support matériel) deux articles consécutifs dans le texte ne sont évidemment pas nécessairement subordonnés au même objet, ni même les objets auxquels ils sont subordonnés ne sont-ils forcément de la même classe (en conséquence de la remarque précédente): ainsi, l’article L131-8 subordonné à la section 2 (chapitre I, titre I du livre I est suivi l’article L132-1 subordonné au chapitre II, dépourvu de section, sous-section et paragraphe (Figure 2). La relation de sélection R_S relie un article qui cite à l’article cité (c’est ce type de relation que nous avons considéré en section 3, un code citant un autre code). La relation de sélection est orientée. Au contraire de la relation d’influence, la relation sélection peut être réflexive. Ainsi la boucle attachée l’article L132-1 en Figure 2 traduit l’expression « mentionnées à l’alinéa précédent ». Cette référence est faite vers un objet de classe intra-article (un alinéa) mais est assignée à l’article dans la Figure 2 dont la résolution se limite à la classe article. Cet exemple illustre en outre le fait que les sommets d’arrivée comme de départ peuvent être des objets de n’importe quelle classe (et pas seulement des articles).

²⁴ Même si certains d’entre eux font explicitement partie d’une classe déjà définie pour les objets internes au CEL (par exemple un article de loi externe sera mis dans la classe des objets externes et non dans la classe des articles).

²⁵ La partie législative du code de l’environnement rassemble les règles de 39 lois préalablement disséminées, ainsi que les autres textes votés après sa première adoption (en 2000).

²⁶ Cette hiérarchisation peut être aussi vue comme expression de la souveraineté accordée à l’autorité qui codifie (et aux autorités dont elle dépend) sur la matière juridique concernée. Aucune institution souveraine n’existe au niveau international qui permettrait de codifier – et donc de décider d’une hiérarchisation des objets du réseau des textes internationaux (sans parler d’une « harmonisation », terme et donc projet lui-même porteur de valeurs spécifiques), qui supposerait résolu l’épineux problème de la préséance des diverses sources du droit.

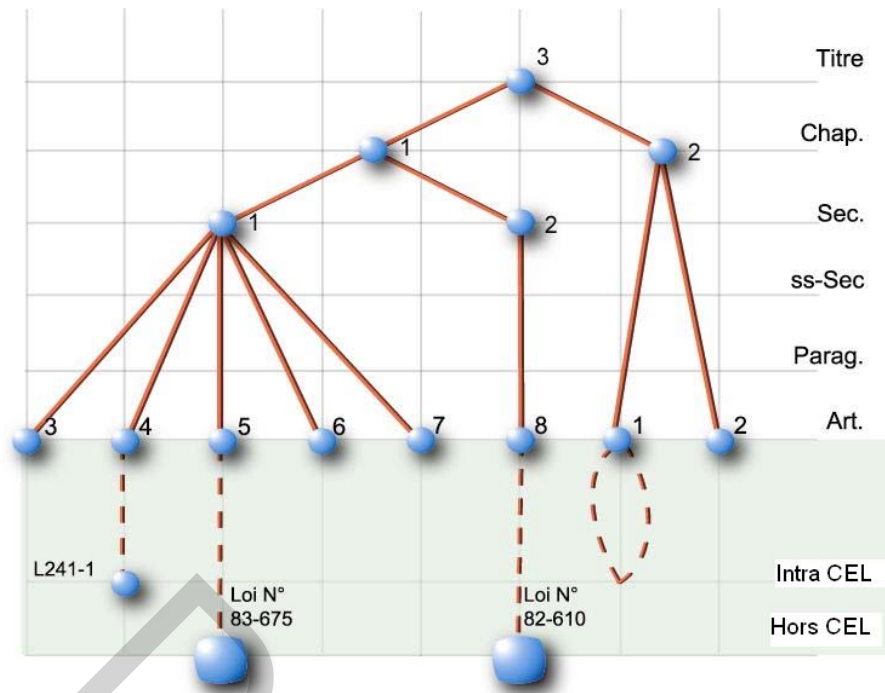


FIGURE 2: Graphe associé au titre III du livre I de la partie législative du Code de l'Environnement (noté *CEL* ici), avec une résolution à la classe "article". Les classes de niveau sont indiquées colonne de droite (de « titre » à « article »). Les relations d'*influence* sont représentées par des arêtes continues, les relations de *sélection* (références) par des arêtes en pointillé. La numérotation des articles se lit de la manière suivante : le n° de livre n'est pas représenté dans cette figure; ensuite viennent le n° de titre, de chapitre puis un identifiant d'article. Ainsi le sommet 8 correspond à l'article L131-8.

4.2 Chemins et distances

Dans notre contexte nous pouvons séparer des caractérisations reposant exclusivement sur la relation d'influence ou sur celle de sélection, soit sur les deux confondues. La notion de chemin nous conduit à celle de graphe connecté: G est connecté si chaque paire de sommets de G appartient à un chemin. Un des objectifs de la codification est de rassembler les textes épars traitant d'une matière donnée en un graphe connecté sous forme d'arbre. La distance maximale observée dans un graphe est son *diamètre*.

Il peut y avoir plusieurs chemins entre deux sommets (notamment dans un réseau). Par exemple un chemin de longueur $l=7$ relie les articles L131-4 et L141-1 du *CEL* : partant de l'article L131-4, on parcourt la séquence de sommets suivants: section 1, chapitre Ier, titre III, livre I puis titre 4, chapitre Ier et article L141-1. Toutes les arêtes de ce chemin traduisent des relations d'influence. Un autre chemin existe, de longueur $l=1$, puisque le L131-4 fait référence au L141-1 (relation de sélection).

On définit la distance entre deux sommets (u, v) comme la longueur du plus court chemin entre u et v . Dans notre exemple, cette distance est égale à 1, via le chemin traduisant la relation de sélection. Ces deux articles sont dits « voisins ». Ainsi les textes directement référencés dans le *CEL* mais externes à celui-ci, sont-ils aussi des textes voisins. Il est possible de suivre un chemin en procédant de voisinage en voisinage, démarche propre à la délimitation d'un corpus traitant d'une matière particulière. Un symptôme de la complexité du droit consiste en l'impossibilité de réunir tous les textes impliqués dans le traitement d'une matière donnée – par exemple l'environnement, ou l'eau – à l'intérieur d'une *unique compilation*, indépendante, c'est-à-dire non connectée à d'autres corpus. En effet, cette sorte

de compilation, utile à l'usage, traverse les frontières qui délimiteraient les sources des textes rassemblés, ou les séparations entre droit interne, communautaire ou international.

Un autre office de la codification, si elle est bien réalisée à moyens légistiques constants, est de produire des codes de *même diamètre*, à savoir, à la résolution de l'article, 16 (et 22 si nous considérons 3 niveaux intra-articles supplémentaires) : le chemin pour remonter d'un article et la racine « Code » est de longueur 8 si tous les niveaux sont représentés (paragraphe, puis sous-section, etc.), et il en faut autant pour redescendre de la racine à un autre article. En considérant uniquement la relation d'influence, la distance minimale entre deux articles est de 2 (par ex. distance entre les articles L131-3 et L131-4, ou entre L132-1 et L132-2 de la Figure 1). Dans la mesure où tous les codes produits suivent la même organisation hiérarchique, leur diamètre est le même. Dans cette estimation, les relations de sélection ne sont pas prises en considération. Suivant cette approche, on peut proposer une *nouvelle fonction* à la codification : celle de transformer les relations de sélection en relations d'influence.

4.3 Le code de l'environnement : régularités statistiques

Nous avons jusqu'à présent décrit le graphe et certaines de ses propriétés en considérant quelques sommets et arêtes. Nous donnons maintenant une description globale du graphe de la partie législative qui nous permettra également de considérer certaines de ses régularités ayant une *pertinence statistique*. Pour cela nous construisons diverses distributions D : il s'agit de compter le nombre (ou la fréquence) d'occurrence d'objets dans une classe déterminée. Certaines de ces distributions empiriques \hat{D} (histogrammes) peuvent également être comparées à des distributions théoriques ou modélisées sur la base de la représentation des processus susceptibles de les générer. L'estimateur \hat{D} diffère généralement de la distribution théorique D (non observable) du fait du nombre limité de données à disposition²⁷, du mélange de plusieurs processus dans la réalisation de la variable aléatoire et enfin d'éventuelles erreurs d'approximations ou de biais dans la procédure de calcul de \hat{D} à partir des données.

G/V_l est le graphe connecté induit par la seule relation d'influence. Il décrit l'organisation hiérarchique du CEL²⁸. En comptant le nombre d'objets par niveau du graphe²⁹ on observe tout d'abord que le nombre d'objets n'est pas une fonction monotone croissante du niveau. En effet, il y a plus de chapitres ou de sections que de sous-sections ou de paragraphes. Cette observation semble indiquer une fonction spécifique aux classes « chapitre » et « section », fonction que ne remplit pas la classe « article ». Contrairement à cette classe, les chapitres et sections identifient, classent, subdivisent la matière traitée par l'opération de codification³⁰, plutôt que le contenu normatif lui-même. Les titres qui les surplombent définissent les grandes lignes de cette organisation, mais ce n'est qu'à partir du

²⁷ Alors que D résulte de théories qui supposent un nombre généralement infini de réalisations de la variable aléatoire.

²⁸ Sachons que le plan de classement ou table des matières est l'objet de discussions pendant tout le processus d'élaboration du code et qu'il ne consiste pas seulement à « classer » suivant une quelconque rationalité des documents dans une arborescence : il a une valeur politique. Il n'est pas neutre, par exemple, de décider que les dispositions concernant la Corse font partie de dispositions particulières par rapport aux dispositions générales.

²⁹ Dans la version de 2006 du CEL nous recensons 3 523 objets (sommets du graphe) distribués comme suit : 7 livres, 30 titres, 115 chapitres, 200 sections, 99 sous-sections, 21 paragraphes, 1266 articles et 1785 subdivisions d'article. Le Code de l'Environnement évoluant, la version de mars 2009 (seule partie législative) présente 31 titres, 122 chapitres, 201 sections, 107 sous-sections, 26 paragraphes et 1288 articles.

³⁰ Matière que nous nommerons parfois « matière régulée » avec l'idée d'aborder plus tard des corpus juridiques anglo-saxons pour lesquels cette expression ferait davantage sens.

chapitre que la matière, objet de droit, exprime sa complexité propre et inaliénable à des classifications juridiques.³¹

Seconde observation : nous recensons à peu près autant d'articles que d'objets ou divisions intra-articles identifiées par un chiffre ou une mention « nota » (nous ne comptons pas ici les alinéas simples). Cette propriété est d'autant plus remarquable que nous considérons deux niveaux de hiérarchie intra-article (chiffres romain plus nota, chiffres arabes). Rappelons que dans une classification dichotomique complète – la moins « inflationniste » possible - il y a 4 fois plus d'objet de niveau $n+2$ que d'objets de niveau n . En terme d'économie d'objets identifiés, il serait tout aussi parcimonieux d'élever au niveau d'articles les objets intra-articles et de supprimer les niveaux correspondants. De fait, de nombreux articles du *CEL* doivent déjà être dépourvus de subdivision (et une seule subdivision n'a pas de sens, à moins qu'elle ne subsiste après abrogation d'au moins une seconde subdivision). A ce stade nous devons donc nous attendre à deux populations bien différentes d'articles : celle, majoritaire, des articles *simples*, monolithiques, et celle des articles *complexes*, pourvus de subdivisions. Reste à comprendre le processus légistique qui induit cette distinction.

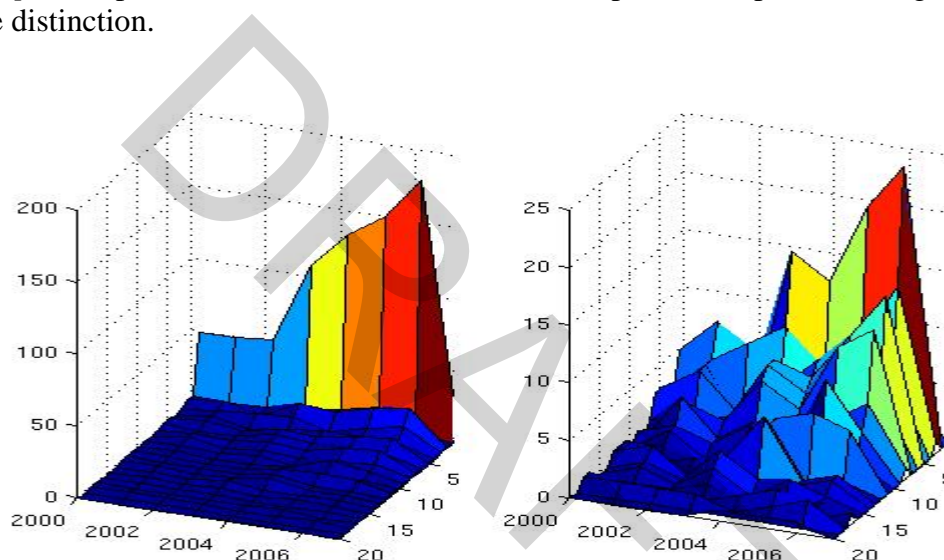


FIGURE 3 : Dans le plan horizontal sont indiqués d'une part l'année de modification des articles (de 2000 à 2007), d'autre part le nombre de subdivisions des articles. Sur le plan vertical on donne le nombre d'articles modifiés du *CEL* correspondant à chacune des classes du plan horizontal. A gauche : tous les articles modifiés sont dénombrés (de 0 à 20 subdivisions) ; à droite, ne sont décomptés que les articles modifiés ayant de 1 à 20 subdivisions.

Peut-on ensuite retrouver dans le document juridique des indices de la dynamique du corpus étudié ? La réponse est positive car les articles du *CEL* qui ont été modifiés doivent indiquer une ou plusieurs dates correspondant à leurs modifications par loi ou ordonnance. Pour chacun d'eux nous avons conservé l'ensemble des dates indiquées et pour chaque année représentée dans notre corpus – à savoir de 2000 au premier trimestre 2007 – nous avons reporté le nombre d'articles en fonction de leur nombre de subdivisions. A nouveau notre échantillon ne porte que sur un nombre assez restreint de dates (1153 dates recensées pour seulement 828 articles. Les résultats sont représentés en Figure 3 (à gauche pour tous les articles, à droite pour le sous-ensemble des articles complexes).

Nous observons entre 2000 et 2006 une augmentation très significative du nombre d'articles modifiés (Fig. 3 gauche). Par ailleurs, le recensement des modifications n'est pas exhaustif pour l'année 2007. En Figure 3 droite où la population des articles simples a été

³¹ Dans d'autres codes, cette hypothèse est vérifiable.

retirée nous observons la même tendance : tous les articles, quel que soit leur degré de subdivision, sont modifiés. La « stabilité » d'aucun type d'article ne semble particulièrement assurée. Les modifications touchent autant les articles simples que les articles fortement structurés, de manière aléatoire pourrait-on dire.

Nous avons ensuite relevé l'ensemble des références trouvées dans le *CEL* qui, d'un point de vue formel, constituent le réseau des relations de sélection. Ces références peuvent viser des articles, mais aussi tout autre objet de niveau hiérarchique, qu'il soit supra (e.g. code, livre, etc.) ou intra-article (alinéas numérotés ou non, etc.). Des textes qui n'appartiennent pas à la hiérarchisation de la codification sont également appelés par référence : lois, décrets, directives, protocoles, conventions et ainsi de suite. Ces références renvoient à des objets appartenant au corpus codifié ou non codifié du droit interne, comme à tout autre texte légal communautaire ou international. Le repérage de toutes ces références pose divers problèmes que nous n'aborderons pas ici : difficultés liées à l'analyse syntaxique (découpage) des textes, à la diversité des formes linguistiques du référencement, au renvoi à des listes ouvertes³², aux citations à l'intérieur d'articles reproduits *in extenso* (l'article dit « suiveur » reprend intégralement le texte de l'article « pilote »), à la présence de références ambiguës, etc. Du fait de ces difficultés, la liste des références repérées dans le *CEL* est nécessairement entachée d'erreurs. Divers tests et vérifications nous conduisent à estimer la précision des pourcentages les plus élevés donnés ci-après à 1 ou 2 %.

Nous trouvons 4739 références, dont 2476 issues d'articles « simples » (non subdivisés) et 2263 issues d'articles *complexes*. Si nous rapportons ces chiffres aux nombres d'articles de chaque classe, nous obtenons en moyenne 2,65 références par article simple contre 6,80 références par article complexe. Ainsi la complexité juridique - que nous cherchons à caractériser - n'est pas seulement gérée en intégrant trois niveaux supplémentaires de subdivisions aux articles³³, mais aussi en faisant de ces divers niveaux des sources denses de références vers divers documents ou corpus eux-mêmes plus ou moins organisés de façon complexe. La complexité se reporte de documents à documents, de corpus à corpus.

Par ailleurs, près de 3500 références visent des objets propres au *CEL* lui-même. Des analyses préliminaires effectuées sur une partie du réseau de dépendances d'articles du *CEL* tendent à montrer qu'un effet « petit monde » se dessine. Cela signifie d'une part qu'il est pourvu d'une assez forte connectivité locale (deux articles auront une probabilité d'être liés qui est plus forte si chacun d'eux est déjà lié à un même troisième article) et d'autre part qu'il existe des chemins courts (en moyenne) entre n'importe quel sommets du réseau. Enfin nous pourrions analyser les capacités de *résilience* de la structuration du *CEL*: ainsi des erreurs de référencement devraient-elles rester, statistiquement, sans incidence notable sur l'exploitabilité du code de l'environnement³⁴. On pourrait utiliser ce *calcul de résilience* pour tester différentes hypothèses sur le travail (empirique) du codificateur autant que sur les contraintes du domaine.

Nous observons d'autre part que plus de 63,1 % des références citent des articles, conformément aux prescriptions de la pensée juridique classique et de la légistique. Mais près de 15,9% des références visent explicitement des subdivisions intra-articles. Le rapport de ces deux pourcentages est un peu inférieur à celui du nombre d'articles simples sur le nombre d'articles complexes. Comme nous l'avons vu précédemment, si la densité de renvois à

³² Du type « article <ref_article> et suivants ».

³³ Alinéas numérotés par des chiffres romains, arabes ou non numérotés.

³⁴ Résultats que nous entreprendrions de confirmer par l'analyse d'autres codes ou d'un corpus plus large. Il doit cependant rester clair que cette affirmation, vraisemblable du point de vue de l'analyse de ce système complexe, est certainement marginale quant à la conduite de dossiers spécifiques pour lesquelles la précision et la rigueur d'une instruction sont essentielles à la bonne marche de l'affaire.

d'autres objets est très élevée à partir des subdivisions d'articles, les articles restent préférentiellement les cibles de références. Près de 9,7% des références visent des objets hors de l'emprise de la codification (textes de lois, directives, textes du droit international, etc.), pourcentage élevé qui pourrait être lié à la dimension internationale de certains objets et enjeux environnementaux. Chapitres (~4,2% des citations) et sections (~1,8%) ne sont référencés (ou presque exclusivement) qu'en tant qu'objets du Code de l'Environnement.

5 Vers une approche dynamique du document juridique

Que peut-on inférer sur l'évolution du document juridique à partir des statistiques construites précédemment ? Sur le type des dynamiques mises en œuvre ? Là encore l'approche proposée ouvre de nouvelles pistes d'étude prometteuses pour des innovations légistiques.

5.1 Organiser la complexité juridique pour mieux comprendre la norme

Pour définir la complexité d'un phénomène, il est nécessaire de s'éloigner du constat de la complication. Sont complexes les systèmes composés de nombreux éléments reliés par des interactions dynamiques, c'est à dire qui *évoluent* dans le temps de façon non linéaire (non causale). Comme leurs effets ou comportements sont *imprédictibles*, on ne peut exercer sur eux un contrôle total³⁵. La nature a engendré la complexité et l'a augmentée systématiquement dans tout le vivant. La question fondamentale des biologistes est celle de cette évolution « naturelle » vers *le plus de complexité*.

En général la complexité, dans son sens non-trivial, est l'apanage des sciences physiques et biologiques. Mais on peut aussi étendre l'étude de ces phénomènes aux sciences sociales. Quand les institutions parlent de complexité, elles ne décrivent pas l'état d'un système mais leur difficulté à le maîtriser. Le droit et la science politique traitent du contrôle et de l'organisation des sociétés. La gouvernance, en refondant la question de l'origine des pouvoirs de contrôle, est liée au *plus de complexité* des systèmes sociaux.

En quoi la complexité peut-elle devenir un nouveau *paradigme pour les sciences, de la physique à la biologie et aux sciences du gouvernement* ? La *théorie des systèmes complexes* reprend comme modélisation les notions fondamentales de la complexité pour décrire, prévoir et simuler des systèmes complexes. L'intérêt de cette approche est que nous sommes partis d'une (nouvelle) définition du corpus juridique, celle d'un graphe interconnecté et que les outils dont nous nous sommes servis s'avèrent parfaitement adaptés à notre hypothèse sur la dynamique des systèmes complexes.

Les niveaux et objets sont reliés entre eux par des relations d'influence et de sélection. Le livre organise la matière (l'eau, l'air, etc.) à l'échelle la plus haute. Le chapitre fait le lien entre la matière et son usage social. A l'intérieur des articles sont immergées des relations fortement dynamiques mais où disparaît la notion d'*échelle*. Or, il est fondamental d'observer les échelles où se situent un terme ou une norme. On pourra voir dans la confusion des échelles une des caractéristiques de la complexité de la lecture du droit.

En résumé, un code n'est pas un livre : c'est un système complexe c'est-à-dire un réseau d'objets hiérarchisés en niveaux qui correspondent à des échelles qui ont des usages et des fonctions spécifiques dans un univers structuré.

³⁵ Voir par exemple Alain Cardon, *La complexité organisée*, Lavoisier, 2005.

5.2 Visualiser les lieux de complexité d'un corpus juridique

Le modèle des graphes a pour objectif de *mesurer* des phénomènes de réseaux : la visualisation des résultats obtenus sur le corpus codifié que nous avons analysé a mis en lumière certains aspects de la complexité juridique.

Les principales conclusions auxquelles ce premier travail nous a conduits peuvent être résumées ainsi :

- S'agissant de la structure du réseau des codes :
 - à ce niveau de granularité, le réseau est très dense, c'est-à-dire qu'il présente une forte connectivité entre les codes via leurs inter-citations (ce qui en fait vraisemblablement un type de réseau ayant des caractères spécifiques au droit);
 - des regroupements de codes sous formes de communautés peuvent être révélées qui constituent des structures sous-jacentes du système juridique français ; ces communautés ont probablement émergé au cours de la longue construction du droit par suite d'une organisation cognitive mais collective (« *distribuée* »), de l'ensemble des domaines juridiques ;
 - nous suspectons que ce type de structuration (forte densité du graphe associé, communautés de codes) existe aussi dans de nombreux autres systèmes juridiques considérés à ce niveau de granularité (ou à un niveau équivalent dans un système non codifié).
- S'agissant de la structure hiérarchique du *CEL* :
 - c'est au niveau des chapitres et des sections que la « matière soumise à régulation »³⁶ s'articule avec les choix de « régulation » (rappelons que le code est créé *ex post* à partir de la matière qui lui préexiste);
 - les divisions intra-articles permettent de gérer diverses formes de complexité (technique, ontologique, sémantique, temporelle etc.) autres que normatives ;
 - il existe 3525 objets, toutes classes confondues (1740 au-dessus du niveau de l'article et 1785 aux niveaux internes à l'article) ce qui montre une grande structuration hiérarchique de ce corpus de 1266 articles.
- S'agissant du réseau des références du *CEL*:
 - les articles restent, comme nous le supposions, les principales cibles des références ; ils sont cités 2321 fois à l'intérieur du code, ce qui est un chiffre important compte tenu du nombre total d'articles (1266) ;
 - cependant les articles subdivisés – les *articles complexes* suivant notre terminologie – sont aussi une source dense de références : ils sont cités 615 fois à l'intérieur du code;
 - cette quantification nous a fait comprendre que le niveau de l'article n'est plus le seul pertinent pour appréhender la norme comme unité de mesure du droit, ce qui signifie que le niveau de l'article est fragmenté en sous-niveaux qui font éclater le couple article/norme.
- S'agissant de l'ensemble du graphe que constitue le Code:
 - la « complexité » juridique est en quelque sorte cachée et organisée à l'intérieur même des articles ; ceux-ci sont distribués en un réseau fortement connecté (« petit monde »), ce qui confère à cette organisation des *distances faibles* entre les articles du code ;
 - la dynamique d'évolution de la taille des articles suggère un processus de croissance exponentielle aléatoire (ce n'est pas parce qu'un article est plus long qu'il est plus

³⁶ Ici l'environnement, entendu en un sens très large qui inclut également les acteurs et les institutions environnementaux.

modifié, ce qui signifie aussi que la longueur n'est pas un critère pertinent pour contrôler la dynamique du Code).

Il pourra nous être objecté que les résultats statistiques impliquant les relations de sélection ne représentent que les régularités d'une organisation arbitraire des articles du *CEL*, régularités issues de *stratégies internes aux jeux d'acteurs de la codification*. On a déjà signalé que ce droit était *transversal* car il réunit des dispositions du droit rural, du droit industriel et du droit de l'urbanisme. Le degré d'arbitraire de cette organisation ne peut être estimé que par le codificateur et le législateur. En outre, l'impératif de complétude de la matière, propre à toute codification, ne vient que renforcer la densité des réseaux que nous avons étudiés dans des univers codifiés. Cette logique a été amplement critiquée mais, comme le signalent Lascoumes et Martin³⁷, « il n'y a pas nécessairement antinomie entre complexité d'une règle ou d'un corps de règles et leur expression codifiée ». Il existe en effet d'autres moyens que la cohérence structurelle pour maîtriser la complexité des règles : l'affichage de méta-règles comme les principes fondamentaux d'un domaine.

6 Conclusion

Les résultats présentés proposent deux niveaux d'analyse de la complexité du droit. Nous avons d'abord construit un graphe qui représente les relations croisées entre les 52 codes français. Le réseau des citations entre codes n'est pas orienté, ni pondéré : nous envisageons de caractériser ces relations (définition, exception, etc.) dans une prochaine étape. Dans un deuxième temps, nous avons analysé la structure d'un code, celui de l'environnement. Nous avons défini la notion d'article complexe, de diamètre du réseau des relations, de densité et de résilience du droit. Nous avons aussi commencé à aborder les aspects dynamiques, évolutifs du système juridique. Il est certain qu'une analyse de la complexité juridique devra considérer les phénomènes évolutifs, tout autant que la dimension sémantique et substantielle³⁸ du droit. Ainsi, face au constat de la complexification du droit, amplement commenté et analysé par les juristes, nous proposons de nouveaux outils d'analyse et de maîtrise – au moins partielle – de cette matière en rapide évolution, ainsi que des moyens de *visualiser* les lieux de complexité des corpus juridiques.

Remerciements

Nos remerciements à Mme Elisabeth Catta, Rapporteuse auprès de la Commission Supérieure de Codification, pour ses nombreux commentaires et conseils au cours de ce travail. Les données légales ont été produites avec le support du Département *Recht, Regierung, Technologien* de Berlin (Centre Marc Bloch & CNRS, Berlin). R. Boulet bénéficie d'une bourse post-doctorale attribuée par l'Institut National des Sciences de l'Univers (INSU, CNRS Paris). Cette recherche est en partie financée par le RTRA Sciences et Techniques de l'Aéronautique et de l'Espace (Toulouse) via le projet MAELIA (<http://www.iaai-maelie.eu>) « Multi-agents for Environmental Norm Impact Assessment ».

³⁷ P. Lascoumes et G. Martin, op. cit. p. 10 <http://reds.msh-paris.fr/publications/revue/html/>

³⁸ Une première approche est proposée en considérant les relations entre ontologie juridique et corpus de texte, celui-ci étant considéré comme un support (au sens des distributions statistiques) de fonctions d'information mutuelle entre concepts juridiques. Cf. Mazzega P., Bourcier D., Pourgine P., Nadah N. and R. Boulet (2009) *A complex system approach : legal knowledge, ontology, information and networks*. In *Approaches to Legal Ontologies*, G. Sartor, P. Casanovas, M. Biasiotti & M. Fernández Barrera eds. Springer, *submitted*.